DOCUMENT RESUME

ED 029 552 HE 000 333

By-Creager, John A.

General Purpose Sampling in the Domain of Higher Education.

American Council on Education, Washington, D.C. Office of Research.
Report No-ACE-RR-Vol-3-No-2-1968
Pub Date 68
Note-32p.
EDRS Price MF-$0.25 HC-$1.70
Descriptors-•Educational Research, •Higher Education, •Research, •Research Methodology, •Sampling, Statistical Surveys
Identifiers-•American Council on Education Office of Research

The experience of the American Council on Education's Cooperative Institutional Research Program indicates that large-scale national surveys in the domain of higher education can be performed with scientific integrity within the constraints of costs, logistics, and technical resources. The purposes of this report are to provide complete and up-to-date documentation of the sampling and weighting procedures currently being used in the Program, and to make available a record of the Program's experience in applying survey sampling procedures in practical situations. Discussions are presented on: the general principles and purposes of survey sampling; definition of the domain and population to be sampled; development of the actual sampling design; weighting procedures used to adjust for disproportionate sampling; estimation, source, and control of errors; and sampling of the total data file for special purposes. Copies of this report may be obtained from The Office of Research, American Council on Education, 1785 Massachusetts Avenue, NW, Washington DC 20036. (JS)

ED029552

ACE
RESEARCH
REPORTS

VOL. 3 • NO. 2 1968

(16)

GENERAL PURPOSE SAMPLING IN THE DOMAIN OF HIGHER EDUCATION

JOHN A. CREAGER

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

OFFICE OF RESEARCH
AMERICAN COUNCIL
ON EDUCATION

HE000 333

American Council on Education

Logan Wilson, President

The American Council on Education, founded in 1918, is a *council* of educational organizations and institutions. Its purpose is to advance education and educational methods through comprehensive voluntary and cooperative action on the part of American educational associations, organizations, and institutions.

The Council's Office of Research was established in 1965 to assume responsibility for conducting research on questions of general concern to higher education. *ACE Research Reports* are designed to expedite communication of the Office's research findings to a limited number of educational researchers and other interested persons.

ACE RESEARCH ADVISORY COMMITTEE

Nicholas Hobbs, *Chairman*
Provost of Vanderbilt University
and Director of the Kennedy
Center, Peabody College

Allan M. Cartter
Chancellor
New York University

John G. Darley
Chairman
Department of Psychology
University of Minnesota

N. L. Gage
Professor of Education and
Psychology
Stanford Center for Research
and Development in Training
Stanford University

Richard C. Gilman
President
Occidental College

Ralph W. Tyler
Director Emeritus
Center for Advanced Study
in the Behavioral Sciences

Dael Wolfle
Executive Officer
American Association for the
Advancement of Science

Additional copies of this *Research Report* (Vol. 3, No. 2, 1968) may be obtained from The Office of Research, American Council on Education, 1785 Massachusetts Avenue, N.W., Washington, D.C. 20036.

Foreword

The purpose of this report is twofold. First, it supplies a more complete and up-to-date documentation of the sampling and weighting procedures currently being used in the Council's Cooperative Institutional Research Program. Second, it makes available to others engaged in survey research in education a record of our experience in applying survey sampling procedures in practical situations where scientific considerations must be applied with respect to both costs and logistic hazards.

The author wishes to thank his colleagues on the Office of Research staff for their many suggestions in this phase of the research program, and for their thoughtful review of earlier drafts of this report. Special thanks are due to Catherine White and to Barbara Blandford for preparing and proofing the final manuscript.

John A. Creager
Research Associate

Table of Contents

		Page
	Foreword	
I.	Introduction	1
II.	General Principles and Purposes of Survey Sampling . .	2
III.	Definition of the Domain and Population to be Sampled	4
IV.	Development of Actual Sampling Design	8
V.	Figure I. Stratification Design for 1968 Survey of Entering Freshmen	10
VI.	Table 1. Correlations of Selectivity with Student Responses to Selected Items from the Student Information Form	16
VII.	Weighting Procedures to Adjust for Disproportionate Sampling	19
VIII.	Estimation, Source, and Control of Errors	21
IX.	Sampling of the Total Data File for Special Purposes .	24
X.	Epilogue	26
XI.	References	29

General Purpose Sampling in the Domain of Higher Education

John A. Creager

American Council on Education

The domain of higher education in the United States exhibits an ever-changing pattern of diversity. This diversity may be seen in administrative and fiscal policies, modes of corporate control, types of programs and services offered, and characteristics of faculties and student bodies. The ever-changing pattern is evidenced in the formation and dissolution of interinstitutional groupings, in the establishment of new institutions, and in the occasional demise of an old one. In such a kaleidoscopic domain, nationwide studies of the higher educational system are expensive and difficult to perform, even with the excellent cooperation given by most institutions. Survey sampling methods provide the most feasible approach, and have become more feasible as data on the full population of institutions have become more readily available. Even in cases where the information is sought from survey units other than institutions (e.g., students or faculty members), such units are appropriately sampled within institutions for both technical and logistic reasons.

The purpose of this paper is to summarize the rationale used, and the experience acquired, in sampling for the Cooperative Institutional Research Program of the American Council on Education. Such information may prove useful to others planning research of general significance to higher education. Although the particular experience acquired in this program would have to be adapted to fit the specific requirements

of another program, we can usefully exemplify the principles of applying sampling theory to a practical survey problem, in which cost and logistics must be considered.

The general nature and purposes of the Cooperative Institutional Research Program have been described by Astin, Panos, and Creager (1966). Each year of the program, initial contact is made with freshmen entering participating institutions; this contact results in extensive data which is used to determine national norms on entering freshmen (1967a, 1967b, 1967c, 1968), and which also serves as input data to implement a longitudinal research design. The students are followed up by mail contact at later points in their academic careers to provide data for studying the impact of the college environment on the students. Because the program is now in its third year of full-scale operation, our experience with the annual survey of entering freshmen includes not only those sampling problems encountered in a given year but also those that arise from temporal changes in the domain of higher education.

General Principles and Purposes of Survey Sampling

The primary goal of a sampling design is to ensure that statistics from the sample either are, or can be adjusted to be, representative of the corresponding parameters of a defined population. Probably the most thorough discussion of survey sampling designs and their applications is that by Hansen, Hurwitz, and Madow (1953). A simpler treatment of the main issues may be found in Peatman (1947). The choice of the population and of data to be acquired depend on the purposes of the

survey: i.e., upon what kinds of information are required about what kinds of units. The development and implementation of an appropriate survey design depend on such considerations as costs, logistic requirements, protection against operating hazards, the kinds of data to be collected, and the amount of relevant information available about the domain. These considerations demand a complex, mixed-strategy design in a survey in which the data are used both to determine norms and to serve as input to a longitudinal research program. The Cooperative Institutional Research Program uses a design which involves the sampling of entering freshmen within institutions, differentially and disproportionately stratified within several subpopulations of the institutional population.

In order to design and execute a sampling procedure, it is necessary to define the population to be sampled and to choose the control variables. The choice of control variables depends on the nature, amount, and reliability of information that is available about them. Since these variables are used to control sampling bias, their importance cannot be underestimated. Until recently, the constraints imposed by cost and logistic considerations were so great as to render such an undertaking as the Cooperative Institutional Research Program as impractical as a complete census of all freshmen. Increased availability of relevant information has changed this. That institutions and students can be studied in a systematic and scientific way, with the flexibility and generality required if such studies are to be useful to the academic community, was shown by Astin (1965b). That the necessary cooperation of institutions in implementing a sample survey of students can be

obtained and that normative information relative to a defined portion of the student population can be derived was shown by Astin and Panos (1966) in their discussion of the pilot study of freshmen entering 65 institutions in 1965. Subsequent experience with three full-scale surveys of more than 300 institutions has provided ample confirmation that sample surveys in the domain of higher education can be executed according to scientifically acceptable standards.

The remainder of the discussion deals with several major issues involved in the sampling design required for an extensive, multipurpose survey designed to obtain a heterogeneous data file. These issues include: (1) the definition of the domain and population to be sampled; (2) the development of actual sampling designs; (3) the weighting procedures used to adjust for disproportionate sampling; (4) the estimation, source, and control of errors; and (5) sampling of the total data file for special purposes.

Definition of the Domain and Population to be Sampled

It is useful to think of the domain of higher education in terms of the institutions providing educational facilities beyond the secondary school level. The "domain," then, consists of all the inputs, outcomes, and intervening events that constitute higher educational processes. Any given survey will necessarily be restricted to certain defined aspects of the domain; these, in turn, determine the population to be sampled and the kinds of data to be acquired. Some examples of populations are institutions, faculty members, students, administrators,

and hierarchically ordered subtypes and combinations of these "sampling units."

The development of the sampling design starts with an enumeration of the eligible population. The complete population of institutions of higher education comprises multiversities, universities, colleges, various kinds of professional schools, junior colleges, and even nonaccredited institutions. A nearly complete listing of the full population is provided by the Higher Education General Information Survey (HEGIS) of the United States Office of Education.[1] This survey includes not only accredited institutions but also those which, though not formally accredited, have their credits accepted by at least three accredited institutions. The number and nature of the institutions not included in HEGIS are not known exactly, but there is reason to believe that some technical institutes and some newly founded institutions are excluded. Although HEGIS lists approximately 100 predominantly Negro institutions, McGrath gives a count of 123 in 1963-64 (1965); a few of these, however, have since closed, merged, or undergone a shift in the racial proportions of the student body. It is reasonable to presume that a definition of the institutional population based on the HEGIS list will be nearly complete, except for a few very small institutions that represent a negligible portion of within-institution sampling units.

---

[1] See Education Directory, Part 3, published each year by the USOE. For more detailed data on each institution, the Reference File and the Opening Fall Enrollment File are especially useful. The Office of Research, American Council on Education, takes this opportunity to express appreciation to the National Center for Educational Statistics for making these files available.

Considerations of costs and logistics may impose further restrictions in the definition of the "eligible" population. In the Cooperative Institutional Research Program, we have imposed two such restrictions: (1) that the institution be functioning at the time of the survey; this restriction eliminates the occasional one which becomes defunct or merges during the planning period, and (2) that the institution have the equivalent of a "freshman" (first college level) class with at least 30 members. This restriction eliminates institutions that require one or more years of undergraduate college-level work for admission to their "first class" and very small institutions, which may grow sufficiently to become part of the "eligible" population in subsequent years of the program. Because available data on opening fall enrollments were not broken down into freshmen vs. other "first-time students" during the first two years of the program, some seminaries and professional schools which have no freshmen were included in the definition of the "eligible" population. Improved reporting procedures have made possible a cleaner definition of the population and a better estimation of the weights required to estimate population parameters.

Temporal changes in the institutional population pose some problems for ongoing programs and longitudinal studies. Even those studies in which a single sample is obtained at a particular point in time may soon become obsolete in a rapidly evolving domain. In defining the institutional population, minor problems occur as a result of occasional mergers and of the establishment or dissolution of institutions. Of somewhat greater concern are the problems encountered in counting and

classifying members of multiversity and state-wide systems and branch campuses of universities. Our general practice has been to follow the U. S. Office of Education's treatment of these problems, counting as separate institutions those for which separate enrollment data records are available and using USOE's classification codes. It should also be noted that some branch campuses of universities have two-year programs, some of them terminal and some of them intended to prepare students for completion of baccalaureate work on the main campus.

Some temporal changes occurring in the domain of higher education concern administrative and fiscal matters. Still others have definite functional implications with respect to the educational process itself and to what happens to students. USOE's continuing efforts to keep abreast of these matters have made it possible to avoid the serious bias in survey design and execution that would result from miscounting or misclassifying institutions. Since large numbers of students may be involved, continuous vigilance is required to ensure appropriate stratification in sampling and in the definition of normative groups for which summary data are computed and reported. Comparability of results across normative groups and across years is enhanced either by adhering closely to a well-defined and widely understood system such as that generated by the U. S. Office of Education or by carefully documenting any departures felt to be required by the purposes and design of a particular survey.

Development of Actual Sampling Designs

In any sampling design, the major control of sampling error is achieved by stratifying the population of institutions along dimensions that are known to represent important functional characteristics of the institutions. Random selection of institutions within different levels of these dimensions thus increases the representativeness of the sample. Although the choice of dimensions is ideally determined by their relevance to control of error, the alternatives are necessarily limited by the information available.

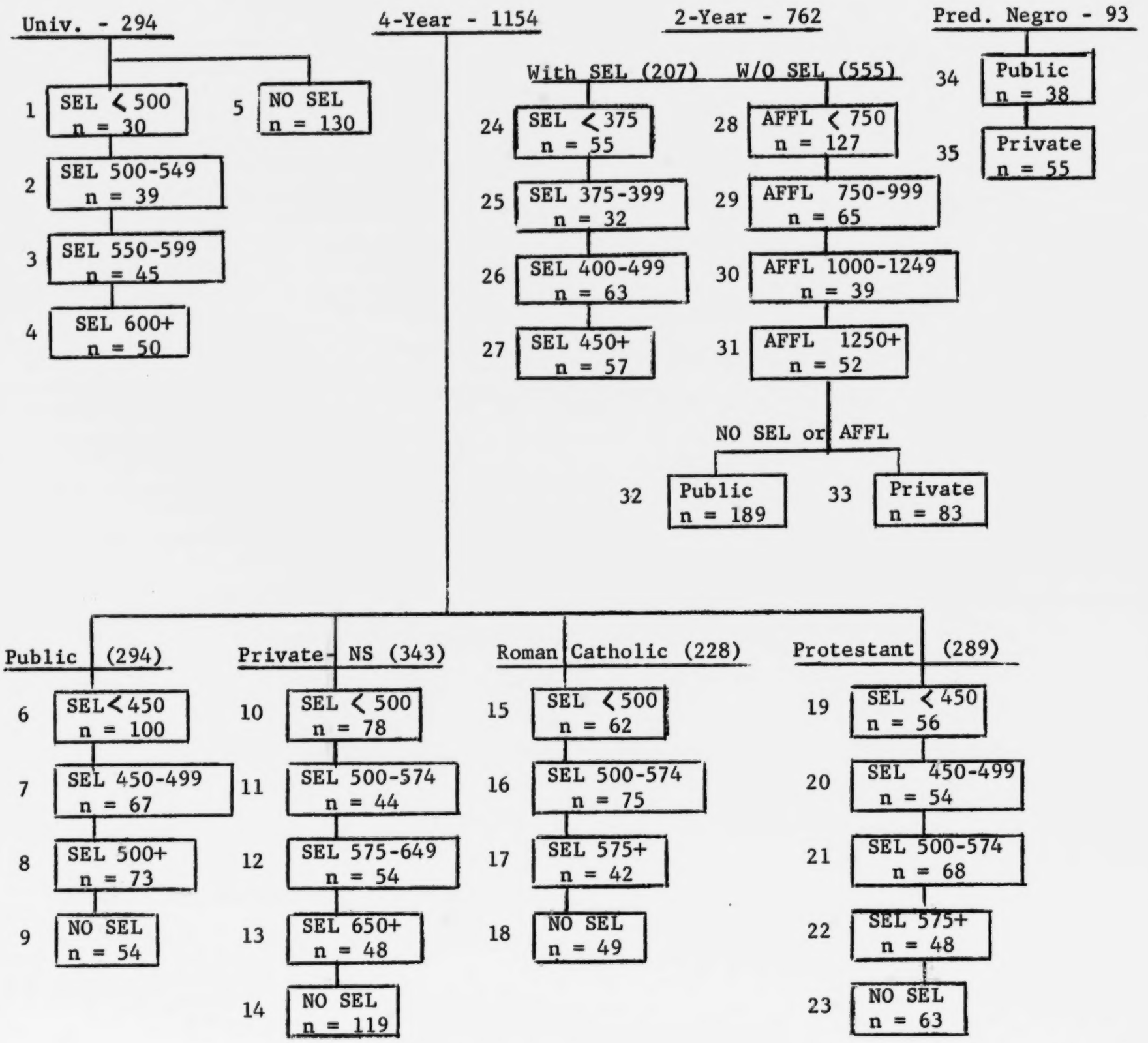
The Cooperative Institutional Research Program uses a mixed strategy in sampling starting with the definition of three subpopulations: universities, four-year colleges, and two-year institutions. This initial division of the population is indicated because these groups of institutions differ widely on a variety of important administrative and educational variables (e.g., size, composition of student bodies, curricula, and college environments). The U. S. Office of Education classification of institutions into these three categories is given in Opening Fall Enrollment in Higher Education, 1967.

The next step in the development of the sampling design consists of stratification on relevant variables within these population divisions, followed by disproportionate random sampling of institutions within the cells defined by the stratification. In the Cooperative Institutional Research Program, the research design involves a wide range of student variables for which no single institutional sampling control variable would be optimal for stratification. For academic

variables such as ability and achievement and for variables highly related to ability (e.g., parents' education, election to high school honor society), some measure of the institution's selectivity--i.e., the intellectual level of its student body--is appropriate. For demographic, personality, and other nonacademic variables, the wide variation in the kinds of students who go to different kinds of colleges (and thus influence and partially define the "college environment") make other institutional characteristics (e.g., size, affluence, mode of control) suitable candidates for control variables. It is a basic principle of stratification that multivariate control quickly reaches a point of diminishing returns in the amount of control of sampling errors for the cost and logistic considerations involved. If there are too many stratification cells, some cells will almost certainly contain too few institutions. On the other hand, insufficient stratification will yield too few, and too heterogeneous cells, with the result that within-cell sampling ratios must be increased to achieve a given level of error control. Just where the balance is to be struck between these extremes is a function of the survey designer's judgment and the resources at his disposal.

Two sampling designs have been used in the Cooperative Institutional Research Program. For the freshman surveys in 1966 and 1967, the three subpopulations were subdivided into a total of 29 cells. In the case of universities and of four-year colleges, the cell structure was based on affluence; in the case of two-year institutions, it was based on size and mode of control. A discussion of the rationale for this approach and of the availability of information on affluence

Figure I. Stratification Design for 1968 Survey of Entering Freshmen*



* SEL and AFFL refer, respectively, to selectivity and affluence as defined in the text.

and size of institution appears in the initial research report from the Program, published by the American Council on Education (1966).[2]

In the third year of the Program, several considerations led to a restratification of the institutional population into 35 cells. This restratification is shown in Figure I. The institutional population has grown rapidly during the last three years, especially the subpopulation of two-year institutions. Not only have new institutions been formed, but also the coverage of existing institutions has improved. A few former two-year institutions have become four-year institutions. More information is available about the various campuses of multi-campus systems. The rapid growth is demonstrated by the increase in numbers of institutions eligible for the survey in each of the three years: 1,968 institutions in 1966; 2,187 in 1967; and 2,303 in 1968.

Past experience in preparing normative information for 24 groups of institutions suggested that sampling errors could be better controlled, especially in the more critical groups, by introducing further breakouts of the four-year institutions (the largest subpopulation). If error

---

[2] Institutional size (total full-time enrollment) and affluence (per student expenditures for educational and general purposes) account for most of the variation among four-year institutions with respect to selectivity, financial characteristics, level of faculty training, and curriculum (Astin, 1962). Affluence and size are also highly related to the college environment (Astin, 1963, 1965a; Astin and Holland, 1961) and to the characteristics of the entering students (Astin, 1965b). Affluence is more strongly related to these other factors than is size. In the case of the two-year institutions, for which affluence data were not then available, the decision to use mode of control and size as stratification variables was based on the research of Richards, Rand, and Rand (1965).

control in the norms groups were the only consideration, a sampling design could be based solely on the norms groups, defined in terms of institutional types. To do so, however, would not allow adequate control of selectivity and affluence. Therefore, both kinds of controls were used in the restratification.

Another development bearing on the decision to restratify is that up-to-date selectivity scores[3] are now available for about two-thirds of the institutions and recent affluence data are available for most accredited institutions including the two-year institutions (Gleazer, 1968; Singletary, 1968). Correlational analyses of relationships among potential stratification control variables, institutional type variables, and variables on which survey data are being obtained have provided further information about the relevance of the stratification control variables used in the survey design.

The primary division of institutions into subpopulations of universities, four-year colleges, and two-year colleges (used in the 1966 and 1967 programs) has been retained. This classification introduces an indirect control on size and some sampling control over about half of the institutional types represented by the various norms groups. The predominantly Negro institutions were separated from these subpop-

---

[3] The selectivity score for each institution is the median standard score on the National Merit Scholarship Qualifying Test taken by those high school juniors in the spring of 1966 who gave the institution as their first college choice (Nichols, 1966). In computing the medians, an adjustment was made for those institutions where the entering enrollments are less than the number of students choosing the college. Both the raw and normalized selectivity scores correlate .91 with the mean SAT Verbal plus Mathematical scores of students actually enrolled by 200 College Board schools in the fall of 1966 (CEEB, 1967).

ulations to form an additional, relatively small subpopulation, in order to ensure better representation and control of this especially interesting group of institutions. Because the predominantly Negro institutions are relatively homogeneous with respect to selectivity and affluence, only two cells, public and private, were formed.

The universities were divided into four cells defined by the distribution of selectivity scores. A residual cell, which contains 130 institutions for which selectivity scores are not available, comprises mainly satellite campuses of public universities, including urban four-year centers, a few two-year campuses, and former state teachers' colleges, often located in small towns. Since affluence scores were available for only ten of them, no suitable basis has been found for further stratification of this heterogeneous "university affiliate" group. The related main campuses appear in the appropriate cells defined by their selectivity scores.

The two-year colleges were first divided into two major subgroups, those with and those without selectivity scores. Those without such scores were stratified on affluence, when data on expenditures were available. This procedure leaves an appreciably large group of schools--most of them relatively new and as yet nonaccredited--for which we have neither selectivity nor affluence scores. The only further breakout made of this group is public versus private.

The large number of four-year institutions permits stratification on mode of control, which also defines some of the norms groups. Therefore this subpopulation was first subdivided into Public, Private-Nonsectarian, Roman Catholic, and All Other Sectarian groups. Within

these subdivisions, there is further stratification based on selectivity scores, with a residual cell in each subdivision for those institutions for which selectivity data were not available. The residual cell in the Private-Nonsectarian group is rather large, but no feasible basis for further subdivision has been found. It should be noted that no stratification control is introduced for the sex composition of the student body, because the production of cell weights and of normative data is done separately by sex.

Within each of the major subpopulations and their subdivisions, alternative cutting points on the selectivity and affluence distributions were examined for possible improvement of cell definition. Selectivity distributions are quite different in the various subdivisions; therefore, the cutting points used to define cells vary from one subdivision to another.

The effectiveness of control variables in reducing sampling error depends on the correlation between the characteristics of the primary sampling unit (institution) and items of information to be collected about the ultimate sampling units (students). The results of two correlational studies were used in designing the stratification procedures for the Cooperative Institutional Research Program.

In the first study, Creager and Astin (1968) examined the interrelationships among 70 variables describing 244 four-year colleges and universities that had participated in an earlier study (Astin, 1965b). Some of these variables proved to be useful in providing direct, indirect, or supplementary controls of sampling errors. For example, the categorical administrative variables are related to size and to those environ-

mental characteristics appearing on a factor determined primarily by size of institution. Selectivity and affluence appear primarily on one large bipolar group factor cutting across factors relating to freshman input, the college environment, and college image.

The second study, specifically designed to guide the restratification, was based on data from 91 four-year institutions in the 1967 survey. In a series of regression analyses, eight of the typological variables, along with selectivity and affluence scores, were correlated with student responses to selected items from the Student Information Form. The results presented in Table 1 confirm the appropriateness of the selectivity score as the primary stratification variable. Selectivity was the most frequent primary predictor: correlations in the .70's are typical with ability and achievement criteria such as high school grades and election to an honor society; correlations in the .40's and .50's are typical with father's occupation, level of family income, student's level of aspiration, and career choice. Selectivity is also moderately related to a wide range of demographic and activity items. These correlations are substantially increased in multiple regression by adding affluence and typological categories. Item types having only a slight relationship with selectivity often have a close relationship with either affluence or the typological variables.

In summary, the restratification improves control of sampling, at little additional cost and logistic effort. It applies a more effective control variable (selectivity), supplemented by highly relevant variables such as affluence and certain U. S. Office of Education type variables used in earlier sampling designs. The more extensive use of

the USOE type variable also provides better control for the sampling of the norms groups defined in terms of these variables. The stratification design can be simplified in studies that involve only certain subdivisions of the population of institutions, a narrower range of types of data, or fewer and more heterogeneous norms groups.

The actual sampling of institutions within cells of the stratification design must anticipate the differential degree of participation by institutions and by students within institutions. The sampling must also anticipate differential loss of data resulting from screening procedures introduced to maintain quality control of the survey data. These hazards can be foreseen in kind--but not exactly in amount--prior to the actual survey. Considerations of costs and logistics, weighed against the desire to minimize sampling errors, led to our decision to obtain a sample of approximately 15 percent of all qualified institutions and as near as possible to 100 percent of the freshmen entering the participating institutions. In view of the possibility that some institutions invited to participate will decline to do so--usually because of difficulties related to scheduling and administering the survey under reasonably uniform conditions--our original planning allowed for 80 percent acceptance from the invited institutions. Each year of the program, the acceptance rate has exceeded this expectation: 82 percent (1966), 88 percent (1967) and 94 percent (1968). That the rate has not only remained high but has even increased may be attributed to two operating policies: useful summary information from the surveys is fed back rapidly, and participating institutions are reinvited. Nonparticipation because of diffi-

Table 1

Correlations of Selectivity with Student Responses to[a] Selected Items from the Student Information Form

Student Information Form Item	Correlation
High School Grade Point Average	
A or A+	.68
A−	.74
B+	.58
B	−.17
B−	−.48
C+	−.70
C	−.65
Father's Education	
Grammar School	−.46
Some High School	−.66
High School Graduate	−.38
Some College	.34
College Graduate	.61
Postgraduate	.52
Father's Occupation	
Doctor	.55
Lawyer	.46
Business	.46
Engineer	.44
Farmer	−.32
Skilled Laborer	−.44
Semi-skilled Laborer	−.55
Unskilled Laborer	−.54
Annual Family Income	
Less than $4000	−.47
4000 - 5999	−.65
6000 - 7999	−.44
8000 - 9999	−.22
10,000 - 14,999	.27
15,000 - 19,999	.61
20,000 - 24,999	.62
25,000 - 29,999	.54
30,000 and above	.54
Elected High School Class President	.51
Won a Varsity Letter	.16
Elected to an Honor Society	.74

[a] Each respondent variable consists of the percentage of students in the institution responding to the item category.

culties in scheduling the survey is not systematically related to the kind of data being obtained and, given the high participation rate, it is unlikely to bias the surveys seriously.

There is, however, another kind of nonparticipation that requires attention in planning and executing a survey: that is, failure to obtain 100 percent of the ultimate sampling units (i.e., the entering freshmen) within each participating institution, can be a source of bias. To guard against this hazard, the institution provides information regarding both the extent of coverage of the freshman class and the quality and conditions of administration. This information is used to judge whether data from a participating institution should be retained as is in the survey, should be retained after adjustment for small random deviations from 100 percent coverage, or should be eliminated entirely with adjustment of the remaining data from the other institutions by appropriate weighting procedures.

A strictly representative stratified random sample would contain a fixed proportion of the institutions in each stratification cell. This procedure was deliberately modified in order to guard against errors resulting from nonparticipation, to reduce the cost per individual student, to protect against accumulating sampling errors in some of the more heterogeneous categories, and to reduce the risk of compounding errors in the aggregate student data. Thus, universities were deliberately oversampled, since the peculiarities of just a few large institutions could introduce an appreciable bias into the student norms. Although including a greater proportion of large institutions increases some of the logistic problems, the risk of peculiarity effects is

diversified over more institutions, with the data from any one institution receiving relatively less weight in the aggregate pooling operations. Institutions in the extreme categories of affluence and selectivity were also oversampled to reduce sampling error arising from the open-ended nature of these categories. Finally, two-year institutions were initially oversampled, since experience indicates that otherwise a sufficient number of participants could not be obtained.

Weighting Procedures to Adjust for Disproportionate Sampling

The data as received from entering freshmen on the Student Information Form constitute a biased sample of the responses of entering freshmen in the defined population: Institutions in the various cells are disproportionately sampled at the time that invitations are sent out; some institutions cannot participate; some participating institutions are unable to obtain a satisfactory sample of their entering freshmen, either because the sample is too small or biased, or because the Student Information Form was administered in such a manner as to cast doubts on the quality of the response data. The first step is to eliminate such questionable data from the survey sample. Fortunately institutional representatives have proved to be not only highly conscientious about quality of administration, but also quite frank about the difficulties they experience. To determine whether the data from a given institution are suitable for inclusion in the normative sample, their reports are carefully studied by the staff.

Student response data retained for inclusion in the survey sample are then weighted, separately for each sex, to make the data reasonably representative of the defined population of entering freshmen. The weights are the product of two factors: one factor corrects for disproportionate representation of institutions and the differential enrollments in the institutions in each stratification cell; the other factor corrects for nonparticipation of students at each institution. The resultant weights are applied to the individual student's data, thus generating normative tabulations for the population of entering freshmen. The weights are also applied to the student response data in the various studies being performed in the longitudinal research program.

In order to obtain the first factor in the student response weights, the entering freshmen enrollments are cumulated across all population institutions in each cell and again for all sample institutions in each cell. The value of the factor, computed separately for each sex, is the ratio of the cell population enrollment to the cell sample enrollment. This major factor in adjusting student response data would be sufficient only if all students in the sample institutions had participated. Therefore, this first factor, based on enrollments in the stratification cells, must be multiplied for each institution by the second factor: the ratio of the freshman enrollment to the number of satisfactory questionnaires returned by

that institution.[4] Application of the student response weight resulting from the product of the two factors is especially important in cross-tabulations and other basic item and composite statistics purporting to be representative of the population of entering freshmen.

When the ultimate sampling unit is the institution and the data either are about institutions or consist of properly weighted averages of student data, institutional weights must be applied to estimate parameters for the population of institutions. Each institutional record is weighted by the ratio of the number of population institutions to the number of sample institutions in the corresponding stratification cell. This procedure permits the stratification design and data files to be used for institutional research as well as for research about students.[5]

Estimation, Source, and Control of Errors

Just how well do the sampling designs function? The statistician

---

[4]Since we usually eliminate those samples which deviate markedly from 100 percent coverage of the freshman class (e.g., 80 percent), this factor is usually very close to 1.00. In the first two years of the program, this second factor was computed for each institutional sample as a whole, without any control for differential participation by sex within the institution. In the third year of the program, we introduced the practice of computing these factors differentially by sex for each institution. It should be noted that the first factor is constant for each college in a given cell of the sampling design, whereas the second factor may vary from one institution to another within a cell.

[5]Institutional decks containing cell numbers, selectivity and affluence scores, are available from the Office of Research, American Council on Education. A general FORTRAN program for generating institutional weights for an arbitrary sample is also available.

refines this question and asks about the "precision" of the sample estimates of population parameters. By "precision" he means how closely the estimates from the sample agree with the value of the population parameter being estimated. In a situation such as the one confronted in the Cooperative Institutional Research Program, this question elaborates into separate questions about the precision of every statistic for each subpopulation or norms group under scrutiny. The complexity of the design and of its implementation make any attempt at formal calculation of precision quite formidable. In a practical situation, however, one need be less concerned with such formal calculations, than with establishing a general picture of the confidence that can be placed in the results as a basis for practical decisions. More specifically, one must establish plausible outer limits for errors of random sampling, judge whether these limits are acceptable, and consider the sources and effects of nonrandom errors from the same practical viewpoint.

Fortunately, the task of dealing with the consequences of random errors can be simplified by considering only categorical percentages (e.g., percent of students choosing a particular item response category). All other statistics can be derived from, and expressed as, combinations (joint and conditional) of what are essentially expressions of item response probabilities. It is also simpler to consider the population of over one million freshmen entering population institutions in a given year as an infinite population, rather than as a finite one, and to ignore the theoretical reduction of standard errors implied by the stratification procedures. These simplifications result in minor

overestimation of the standard errors of categorical percentages. The total normative sample is based on approximately 200,000 respondents. On this basis, the standard error of a categorical percentage of 50 percent as the population parameter is about 0.1 percent. The standard error is theoretically smaller for percentages markedly different from 50 percent. It is larger within the norms groups, based on various subsamples representing subpopulations in the domain. For the smallest norms groups, the standard error may rise to nearly 2 percent.

The chief source of error in stratified sampling is the failure to obtain a truly random sample within each stratification cell. Even though quality control screening and weighting procedures are employed, one must be constantly alert in order to identify and control nonrandom bias. In the absence of knowledge of the true population parameters, it is impossible to ascertain how well such strategy and logistics actually protect against various hazards. However, certain checks indicate that our normative data are well within acceptable limits; these include: (1) consistency in patterns of differences in categorical percentages across norms groups and program years; (2) plausibility of percentages and of distributions defined by an ordered set of categories; and (3) general agreement between our estimates of institutional and student counts and other published data.

To date we have been unable to discover any findings which were wildly out of line, even though one might expect a few simply as a result of random sampling errors and the computation of tens of thousands of categorical percentages. The few small inconsistencies that we have found may be regarded as within two standard errors under random

sampling. It may be argued that no practical decision involving either students or institutions is likely to be affected if a reported categorical percentage used in the decision making is, for example, 25.6 percent instead of 27.4 percent, or vice versa. Some caution may be indicated where such figures are converted into frequencies for estimating loads and facility requirements in some program planning, but even here one is unlikely to obtain more accurate estimates. In the absence of definite information to the contrary, it is reasonable to believe in the scientific integrity of the surveys and of the normative data they produce. Nevertheless, we are open to practical suggestions for evaluating and improving the surveys of entering freshmen.

Sampling of the Total Data File for Special Purposes

In a given program year, the total data file for approximately 300,000 students is processed by computer to create special files for research purposes. These special files include:

1. A 200K (200,000 cases) master file of the students in the normative sample.

2. A 60K random sample file of the normative sample for follow-up studies in the longitudinal research program.

3. A self-weighted 10K file of normative sample students for distributional and correlational analyses.

The 200K master file, which is unweighted, is the basic source file for the creation of additional special files as required in the longitudinal research program. The 60K file is created to reduce the costs of data processing and mail follow-up, costs which might be pro-

hibitive if longitudinal studies were done on the full 200K group. Additional weighting procedures will doubtless be required to correct the follow-up data for appreciable nonresponse to the follow-up surveys, but such procedures may be suitably postponed until that point in the research program.

The decision to select a self-weighted 10K sample for distributional and correlational analyses circumvents the repeated weighting of student response data for each particular study. To form this file, a random sample of students is taken from within each survey sample institution. The number of students sampled is determined, separately for each sex, from the numbers of students in each institution, the size of the total sample to be selected (10K), and the student response weight used in arriving at national norms for entering freshmen. Both the 1966 and 1967 self-weighted samples have been checked against the corresponding national norms. The distribution of deviations in the categorical percentages in the self-weighted sample from those in the national norms is consistent in each year with chance expectations. It is therefore possible to generalize cross-tabulations, distribution parameters, and results of correlational analyses performed on the 10K sample to the defined population of entering freshmen; the result is considerable reduction of processing costs with only a slight loss of precision.[6]

---

[6] The precision can be roughly estimated by treating the 10K file as a random sample of an infinite population. On this basis, the standard error of a categorical percentage is about 0.5 percent for a population parameter of 50 percent.

Epilogue

Large-scale national surveys in the domain of higher education are no longer merely a theoretical possibility. They can be performed with scientific integrity within the constraints of costs, logistics, and technical resources. Any of these constraints, if sufficiently severe, preclude large-scale surveys, but they are no longer insurmountable. It should be noted that no amount of care in the design and execution of the research program or in its procedures for sampling will compensate for poor item sampling or other badly designed features of the survey instrument. Here too constraints which limit the resolving power of the survey instrument may exist. For example, limits of testing time and processing costs require that the range of information obtained from a self-administering questionnaire be maximized. With a volume of 300,000 respondents per year, the hand scoring, tabulating, and punching of item responses may prove formidable tasks, but any marked reduction in this volume would seriously limit the precision and analytical flexibility of the data. The availability of modern optical scanning and document reading equipment, which directly outputs the information onto computer tapes, solves this problem, provided that the system used is itself flexible and accurate and that quality control checks are incorporated at every step of the processing.[7]

---

[7] The survey data in the Cooperative Institutional Research Program have been processed by National Computer Systems, Inc. of Minneapolis, Minn.

Educational research workers have, then, the capability of performing such large-scale surveys. The design of surveys on a more moderate scale may be improved within the constraints faced by the individual researcher, who must adapt the design considerations discussed in this paper to his own particular needs. In most cases, he will be dealing with a simpler situation, sampling only certain subpopulations of institutions, collecting data about institutions rather than about students, or collecting a narrower range of data (perhaps in greater depth). He may have more or less money, staff resources, data processing capability, and cooperation from his sampling units. All of these factors, as well as the scientific considerations emphasized here, will inform his decisions about survey design. It is to be hoped that he will make available information about his experiences, not only to document the quality of his own work, but also to permit colleagues to benefit from his thought and experience.

References

Astin, Alexander W. "An Empirical Characterization of Higher Educational Institutions." Journal of Educational Psychology 53: 224-235; October 1962.

_____. "Further Validation of the Environmental Assessment Technique." Journal of Educational Psychology 54: 217-226; August 1963.

_____. "The Inventory of College Activities (ICA): Assessing the College Environment through Observable Events." Mimeo. Washington, D.C.: American Council on Education, 1965. (a)

_____. Who Goes Where to College? Chicago: Science Research Associates, 1965. (b)

_____; and Holland, J. L. "The Environmental Assessment Technique: A Way to Measure College Environments." Journal of Educational Psychology 52: 308-316; 1961.

_____; and Panos, Robert J. "A National Research Data Bank for Higher Education." Educational Record 47: 5-17; 1966.

_____; _____; and Creager, John A. "A Program of Longitudinal Research in Higher Education." ACE Research Reports, Vol. 1, No. 1. American Council on Education, 1966.

_____; _____; and _____. "National Norms for Entering College Freshmen--Fall 1966." ACE Research Reports, Vol. 2, No. 1. American Council on Education, 1967. (a)

_____; _____; and _____. "Supplementary National Norms for Freshmen Entering College in 1966." ACE Research Reports, Vol. 2, No. 3. American Council on Education, 1967. (b)

_____; _____; and _____. "National Norms for Entering College Freshmen--Fall 1967." ACE Research Reports, Vol. 2, No. 7. American Council on Education, 1967. (c)

College Entrance Examination Board. Manual of Freshman Class Profiles, 1967-69. New York: the Examination Board, 1967.

Creager, John A.; and Astin, Alexander W. "Alternative Methods of Describing Characteristics of Colleges and Universities." Educational and Psychological Measurement 28: 719-734; Autumn, 1968.

_____; _____; Boruch, Robert F.; and Bayer, Alan E. "National Norms for Entering College Freshmen--Fall 1968." ACE Research Reports, Vol. 3, No. 1. American Council on Education, 1968.

Gleazer, Edmund J., Jr., editor. American Junior Colleges. 7th edition. Washington, D.C.: American Council on Education, 1968.

References

Hansen, Morris; Hurwitz, William N.; and Madow, William W. Sample Survey Methods and Theory, Vol. I. New York: John Wiley and Sons, 1953.

McGrath, Earl J. Predominantly Negro Colleges and Universities in Transition. New York: Columbia University Press, 1965.

Nichols, R. C. "Colleges Preferences of Eleventh Grade Students." NMSC Research Reports, Vol. 2, No. 9. National Merit Scholarship Corporation, 1966.

Peatman, John G. Descriptive and Sampling Statistics. New York: Harper and Brothers, 1947.

Richards, J. M., Jr.; Rand, Lorraine M.; and Rand, L. P. "A Description of Junior Colleges." ACT Research Reports, No. 5. American College Testing Program, July 1965.

Singletary, Otis, editor. American Universities and Colleges. 10th edition. Washington, D.C.: American Council on Education, 1968.

Other *Research Reports* by staff of the Office of Research:

A Program of Longitudinal Research on the Higher Educational System
Alexander W. Astin, Robert J. Panos, John A. Creager
Volume 1, No. 1, 1966

Some Characteristics of Junior College Students
Robert J. Panos
Volume 1, No. 2, 1966 (out of print)

Evaluation and Selection in the 1966-67 Academic Administration Internship Program
John A. Creager
Volume 1, No. 3, 1966

Trends in the Characteristics of Entering College Students, 1961-1965
Alexander W. Astin
Volume 1, No. 4, 1966

**National Norms for Entering College Freshmen–Fall 1966*
Alexander W. Astin, Robert J. Panos, John A. Creager
Volume 2, No. 1, 1967 ($2.00)

The Use of Publication Citations in Educational Research
John A. Creager
Volume 2, No. 2, 1967

**Supplementary National Norms for Freshmen Entering College in 1966*
Alexander W. Astin, Robert J. Panos, John A. Creager
Volume 2, No. 3, 1967 ($1.00)

Attrition Among College Students
Robert J. Panos, Alexander W. Astin
Volume 2, No. 4, 1967

They Went to College: A Descriptive Summary of the Class of 1965
Robert J. Panos, Alexander W. Astin
Volume 2, No. 5, 1967

Implications of a Program of Research on Student Development in Higher Education
Alexander W. Astin, Robert J. Panos, John A. Creager
Volume 2, No. 6, 1967

National Norms for Entering College Freshmen–Fall 1967
Robert J. Panos, Alexander W. Astin, John A. Creager
Volume 2, No. 7, 1967 ($2.50)

National Norms for Entering College Freshmen–Fall 1968
John A. Creager, Alexander W. Astin, Robert F. Boruch, Alan E. Bayer
Volume 3, No. 1, 1968 ($3.00)

*The complete national norms reports for 1966 can be obtained from The Publications Division, American Council on Education, 1785 Massachusetts Avenue, N.W., Washington, D.C. 20036 for $3.00 a set.